

Building up a Digital Library with Greenstone

A Self-Instructional Guide for Beginner's

K RAJASEKHARAN
K M NAFALA



Library & Information Division
Kerala Institute of Local Administration
Mulagunnathu Kavu – 680 581, Thrissur, India

Email: librariankila@hotmail.com

Phone: 91-487-2204097

Preface

The purpose of this manual is to serve as a self-instructional guide to those who want to build greenstone collections with Greenstone digital library software 2.72 independently.

The manual can be used as a training hand-out for exposure training that long for one or two days. The basic approach of the manual is to include the essential and easy-to-learn things and to exclude everything else that cannot be learned by new users.

In exposure training, please refrain from showering a lot of information of complex nature to the new participants who cannot digest them all. That many dispel some of them from greenstone.

Practice reminds that learning happens in the form of layers one over the other. When the first layer is not stable, you cannot build up the subsequent layers. This manual deals with the first layer of knowledge on implementation of greenstone. Advance features are not dealt with in this manual.

We express our gratitude to Dr. John Rose for the enthusiasm he hds shown in clarifying our doubts, helping us learn newer changes and suggesting corrections in the draft, without which the document would have more infirmities. Some others have also pointed out a few corrections. Thanks to them.

*Mulakunnathukavu
20-02 -2007*

*K. Rajasekharan
K.M Nafala*

Building up a Digital Library with Greenstone

A Self-Instructional Guide for Beginner's

K Rajasekharan* & K M Nafala**

1. Introduction

A digital library is an organized collection of digital documents, in any format such as text, image, audio or video, that would permit easy access by information users and proper maintenance by librarians. Digital libraries do replace the concept of a library as a repository of knowledge with the notion of it as a provider of information services to users. In this age of exponential growth of information materials in print as well as digital form, both physical and digital libraries are essential and can work in unison, supplementing each other.

The *Greenstone Digital Library Software (GSDL)* offers exciting ways to build and distribute digital document collections. It helps us to publish digital collections on the Internet or on CD-ROM. Within a few minutes time, one can build full-text *search indexes* and *browsing classifiers* for any collection of digital documents. Once initiated, the collection building process will take place mechanically, running into several hours or days for a very large collection.

Downloading digital documents from World Wide Web, organizing them into focused collections and making the materials accessible to others can be a prime application area of digital libraries.

* Librarian, Kerala Institute of Local Administration, Mulagunnathu Kavu – 680 581

** Library Computer Operator, Kerala Institute of Local Administration, Mulagunnathu Kavu –680 581

1.1 Objective of this Guide

The objective of this write up is to narrate, with screenshots, the most essential basic steps in building up a digital library of collection containing a few documents in an easy-to-learn style, in windows operating system.

1.2 Software Installation

Software Download Sites

The GSDL 2.72 software can be downloaded from <http://www.greenstone.org/> or http://greenstonesupport.iimk.ac.in/our_mission.htm.

Java2 Runtime Environment can be downloaded from <http://java.sun.com/>

ImageMagick can be had from <http://www.imagemagick.org/>

Ghostscript can be obtained from <http://www.cs.wisc.edu/~ghost/>

Installation of Software

Install **Java2 Runtime Environment** in your computer as a pre-requisite for installing GSDL software. Then install the **GSDL 2.72 (windows version)** software in your computer. Choose the Local Library mode for installation. You may install **ImageMagick** and **Ghostscript** softwares, required to build image collections and to do advanced conversion of PDF and Postscript documents, respectively. While installing the softwares, just choose the default options in the installation wizard.

2. Building up a Collection with GLI

The simplest way to build a new digital library collection is to use *Greenstone's Librarian Interface*¹ (GLI), a component of Greenstone Digital Library software. GLI allows one to collect the documents, import or assign metadata, and build the documents into a digital library, and convert it into a CD-ROM library.

GLI can be used to perform the following basic activities, while building up a collection: -

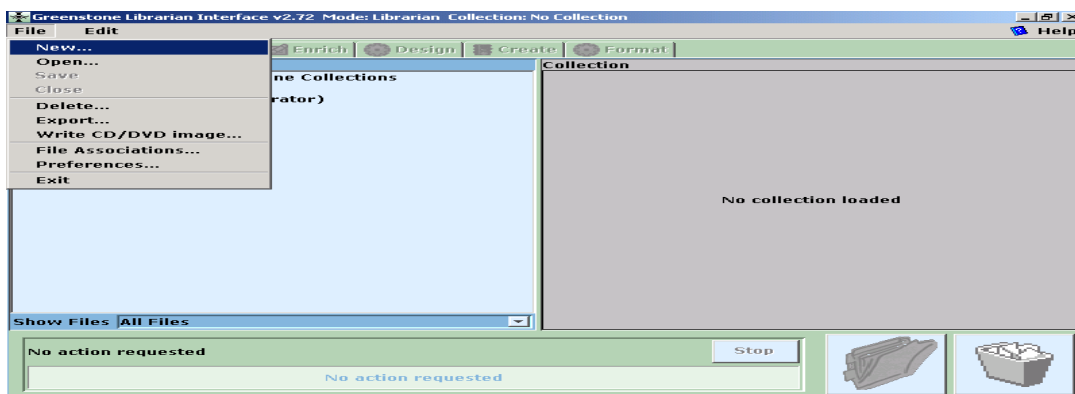
1. **Gather** documents for building up the collection
2. **Enrich** the documents by adding metadata
3. **Design** the collection, its appearance and the access facilities
4. **Format** the appearance of the digital library
5. **Create** the collection
6. **Convert** the digital library into a CD-ROM library

Start a New Collection

Open the GLI from the **Start** ⇒ **Programs** ⇒ **Greenstone Digital Library Software v2.72** ⇒ **Greenstone Librarian Interface**.

To start a new collection, choose **New** from the **File** menu.

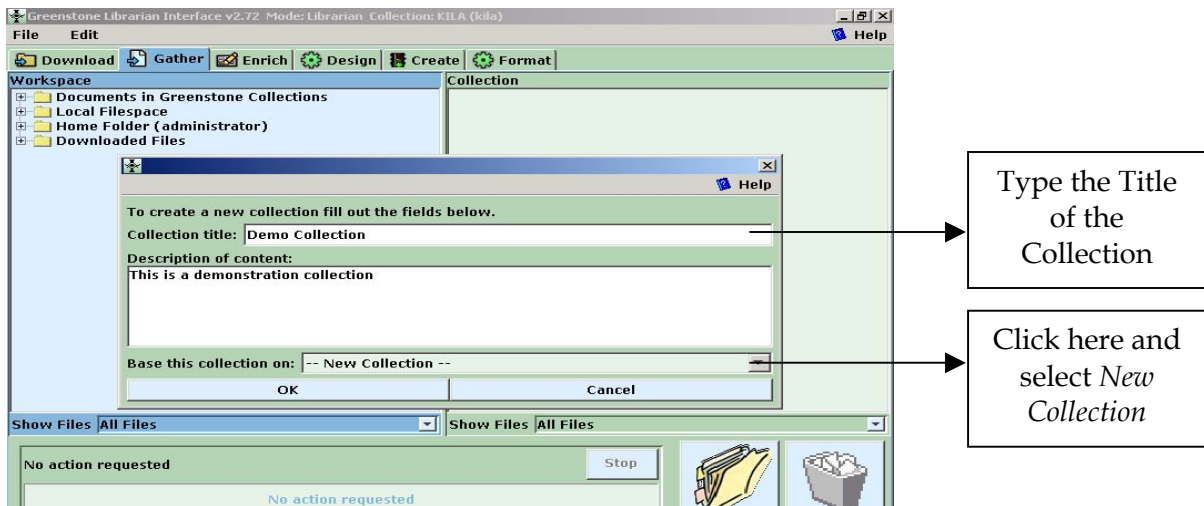
Figure 1 Starting a new collection



¹ GLI is a Graphical Interface program created for making the collection building process easier for librarians who do not have much knowledge to configure a collection. It works in four modes - Library Assistant, Librarian, Library System Specialist and Expert. Librarian is the suitable choice for beginner's and is the default mode. When you go up to the level of Expert, you will get more options to build. You can go to file ⇒ preference to change the mode.

Fill up a name for the collection (against *Collection title*) and a brief description about the collection (against *Description of content*) in the appropriate column in the pop-up window.

Fig 2 Adding Information About the Collection



Choose **New Collection**² in the **Base this collection on** dialogue box and click **OK**.

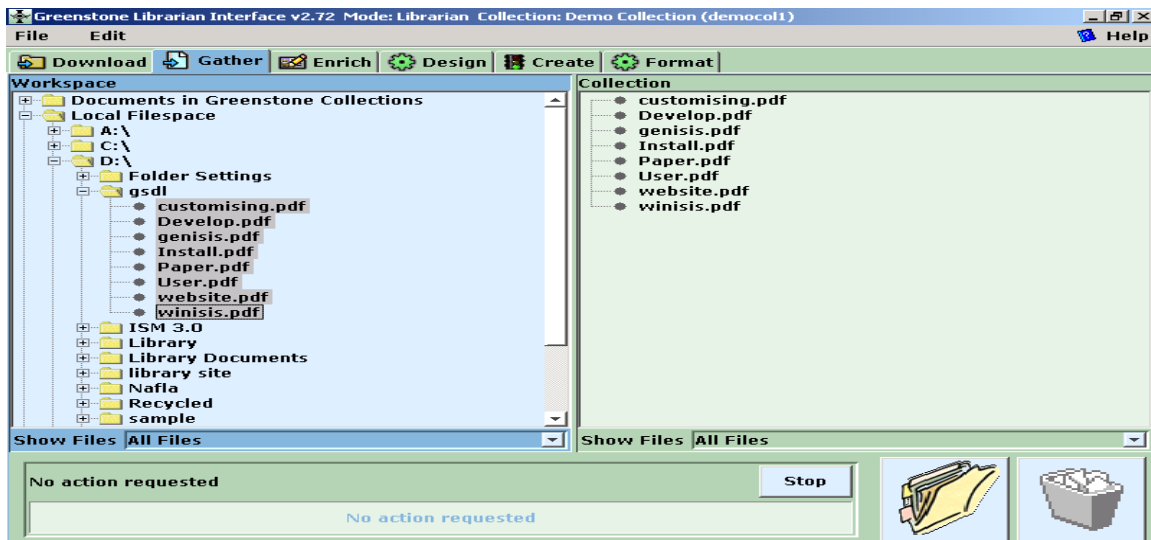
2.1 Gather the Documents

Now the **Gather** panel will become active and it allows the user to collect the required documents by exploring the entire computer. Then select the files or directories by browsing the folders in the computer. Drag and drop them into the right hand pane by your mouse. You can drag the documents either individually or as sets of documents in folders/subfolders.

² This indicates that your collection to be built now will have as entirely new structure. If you want your present collection to follow the pattern of an already existing collection, select the name of that collection from the drop-down menu.

When you gather the documents, the software usually prompts you to select the Plug-in, if the suitable **Plug-in** is not included. In such cases, please click the **Add Plug-in** button.

Figure 3 Exploring the Computer for Gathering the Documents



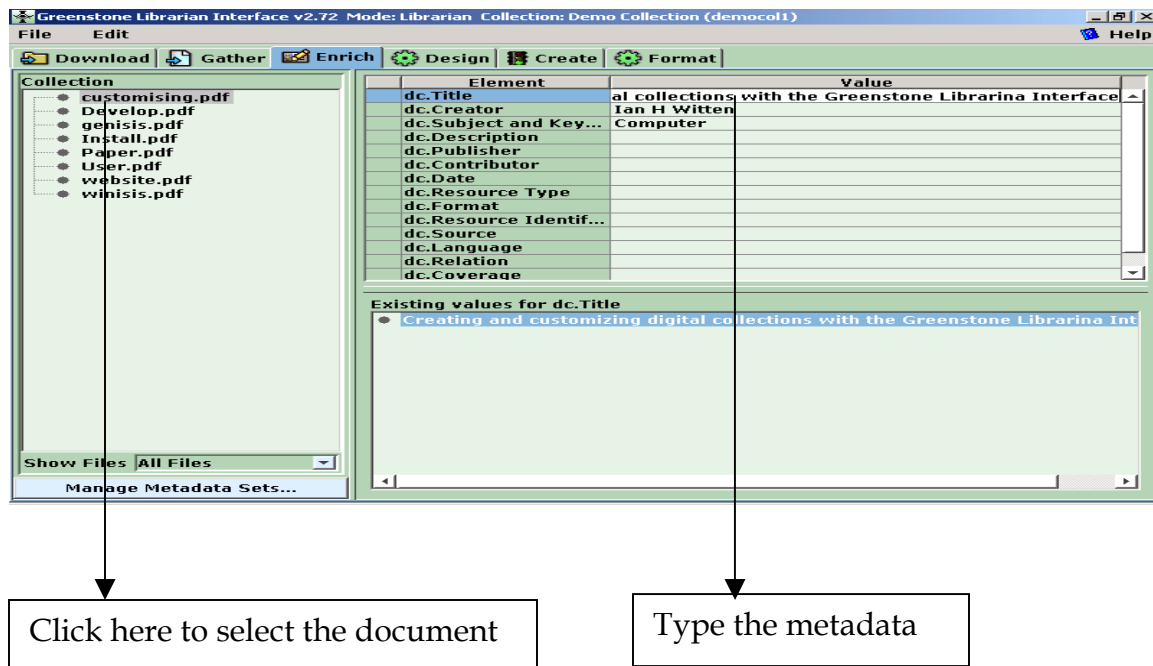
2.2 Enriching Documents with Metadata

The next stage is to enrich the documents by adding metadata³. You can Select the individual document and add metadata such as title, creators or subjects manually.

Click on **Enrich** tab and it will bring up a panel. Left side of the panel under *Collection* tab shows the files. The right side, on clicking, will allow adding metadata for each document on each metadata field in the *Value* box against the *Element*. Here we use Dublin Core metadata and that is why *dc.* is prefixed with name of metadata elements such as Title, Publisher etc.

³ **Metadata** is the data about the documents, such as title, creator, subjects and so on. The metadata element pre-fixed with *dc.* (eg. *dc.Title*) denotes dublin core metadata, *ex.* (eg. *ex.Title*) denotes extracted metadata, *exp.* (eg. *exp.Title*) denotes exploded metadata. Here we use Dublin Core metadata by giving *value* against the metadata element after selecting each document file name under *collection*.

Figure 4 Assigning Metadata Using the Enrich View



Type the Title of the document against **dc.Title**, type the Creator (Author) of the documents against **dc.Creator**, and type the Subjects (Keywords) against **dc.Subject and Keywords** for each selected document shown below *Collection* tab as in the above figure.

2.3 Design the Collection

Then design the collection by choosing the needed features given under the **Design** menu. Collection design consists of many facets as given in the left side pane.

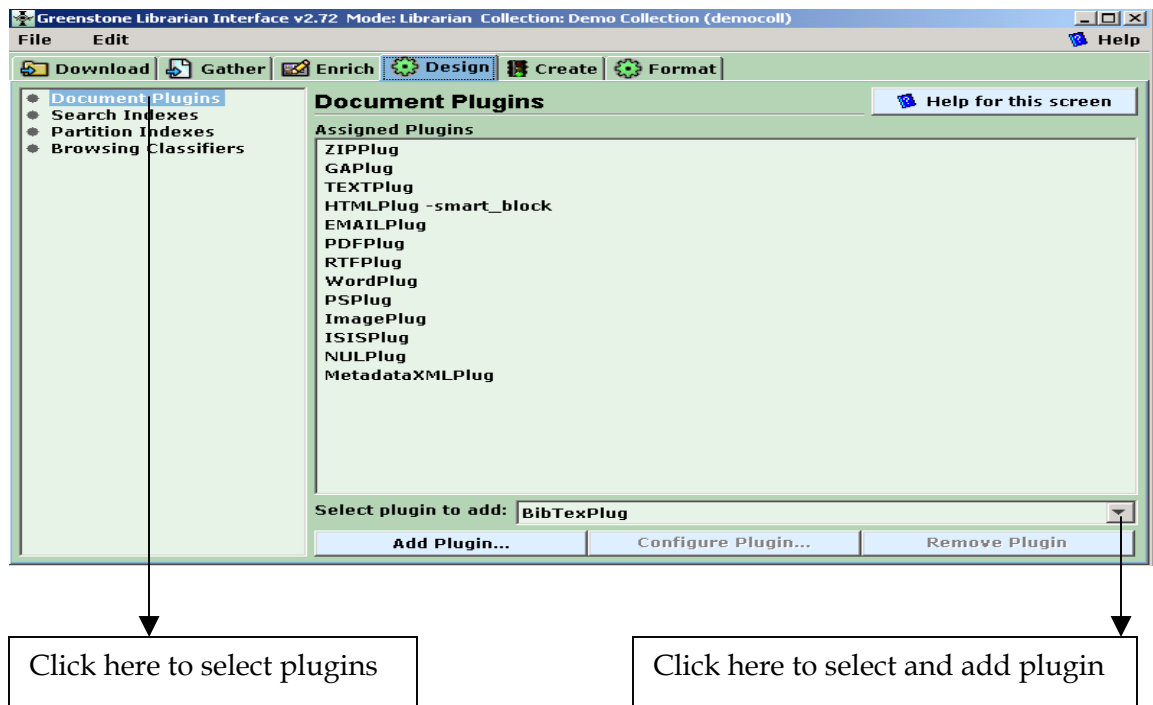
2.3.1 Document Plug-ins

Click on **Document Plug-ins** to add the required Plug-ins⁴ needed to convert the document into the document format (greenstone archive format) required for greenstone. All plug-ins, needed for handling common documents, will be

⁴ Plug-ins, written in perl language, will translate the source document into a common form, parse them and extract metadata from them. For example, the html plug-in converts the Web pages to greenstone archive format and extract metadata that is explicit in the original document from the html source tags.

loaded by default at the time of installation. Kindly note that if proper Plug-in is not loaded, the software cannot build the digital library collection.

Figure 5 Document Plug-ins



2.3.2 Create Search Indexes

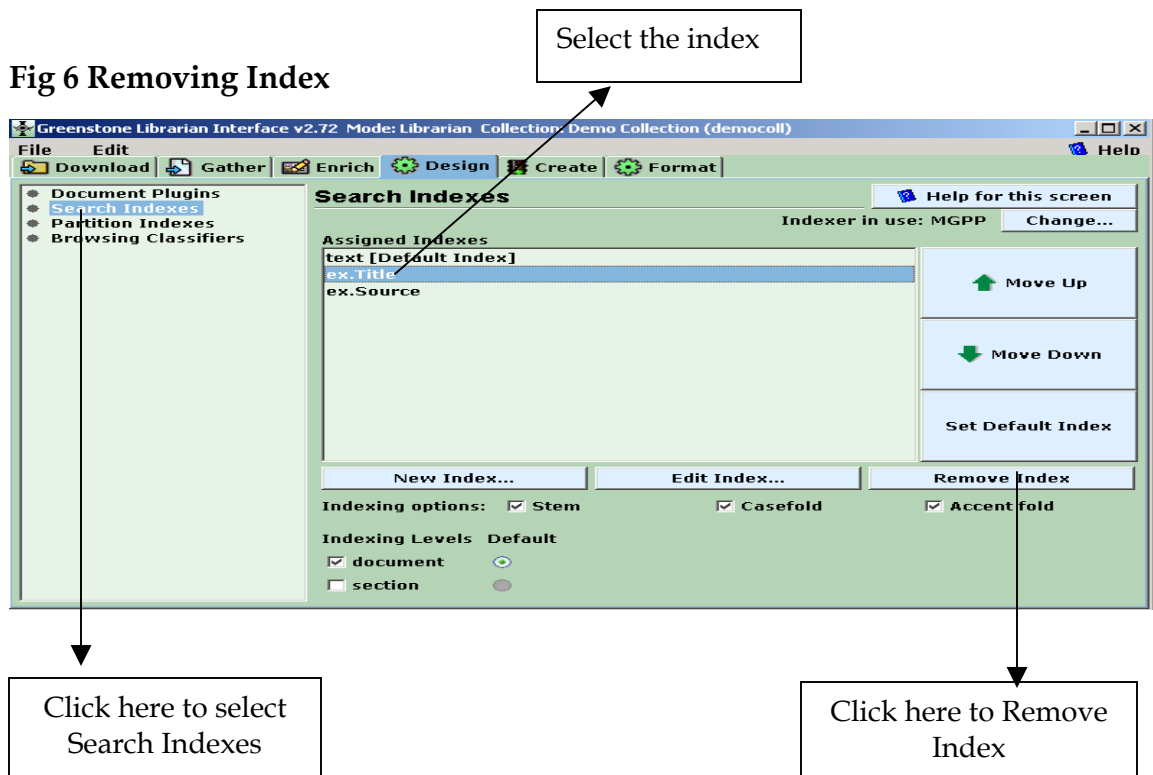
Choose the **Search Indexes**, shown next below on the left pane, for creating *Search Indexes*. Search Indexes determine whether to confine the search to paragraph, chapter or the entire text of the document.

Remove Default Indexes

Remove the default indexes for *ex.Title* and *ex.Source* by selecting the index description under *Assigned Indexes* and then by clicking on the **Remove Index** button. Do not remove the *search index* for **text [Default Index]**⁵.

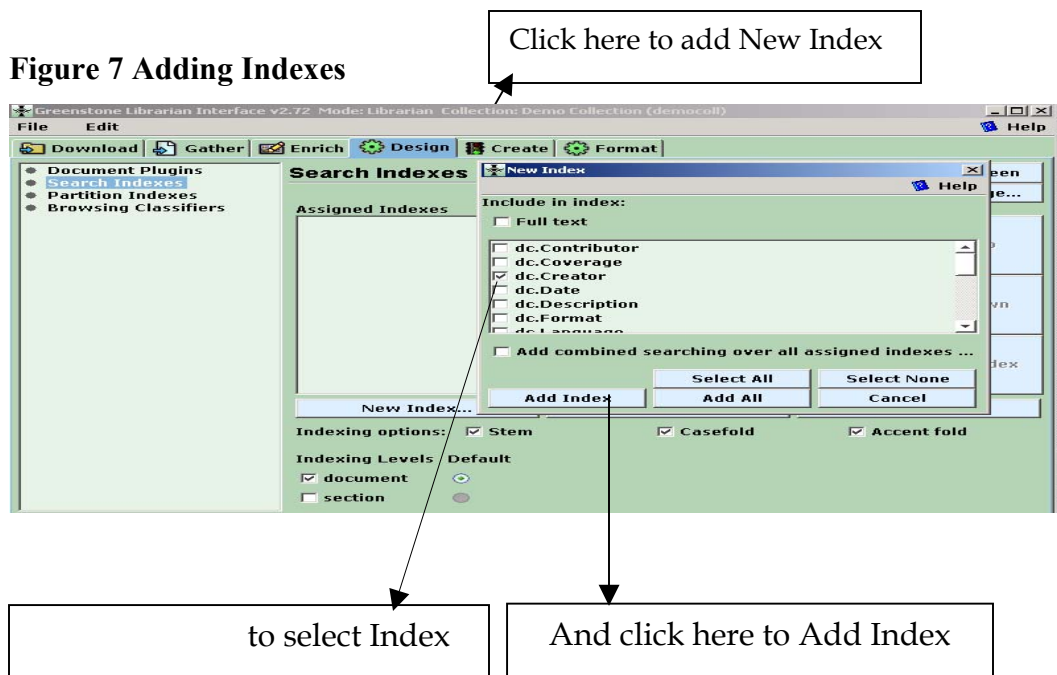
⁵ If you don't provide the search index for text, you cannot search the entire text of the document. Make it default, preferably.

Fig 6 Removing Index



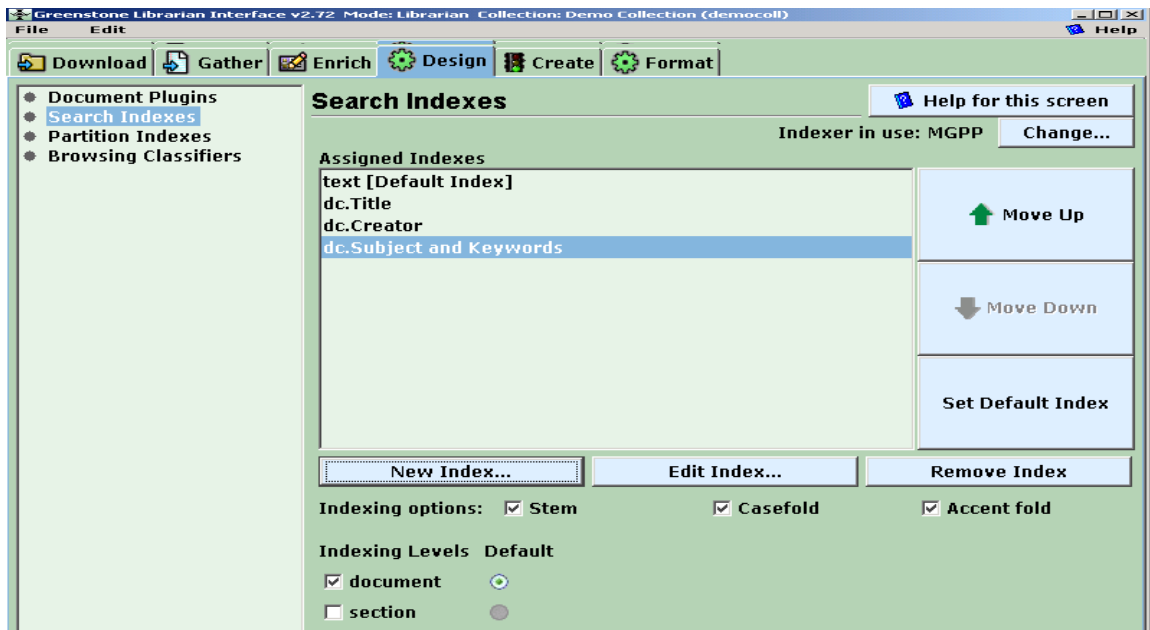
Adding New Indexes

Click on the **New Index** button. Select the **dc.Title**, **dc.Creator** and **dc.Subject and Keywords**, by tick marking on the check box one by one, and add them one by one by clicking on the **Add Index** button. That means select *dc.Title* first and add it, then select *dc.Creator* and add it and so on.



At the end, all the three indexes will be added one by one as in the following figure.

Fig 8 Indexes Selected for the Collection



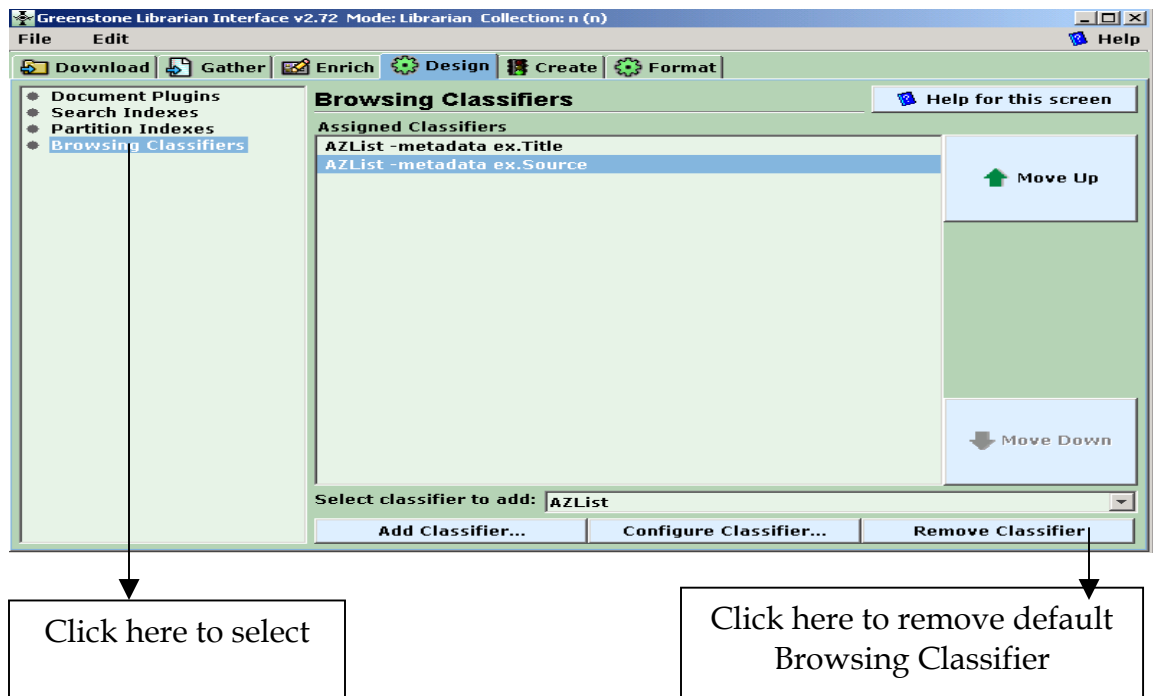
You may select an index and move it up or move it down by clicking on the buttons on the right side so as to set the order of its appearance. Likewise you can set any *index* as *default index* by using the **Set Default Index** button.

2.3.3 Browsing Classifiers

If you want to browse on a metadata element, you must set up a Browsing Classifier, independently of creating an index on this metadata element. Browsing Classifiers such as Titles, Creators, Subjects (see the Fig.20) help you to browse the collection.

Select **Browsing Classifier** by clicking on it and then click on the **Remove classifier** button for removing them one by one. Remove the default Browsing Classifiers for **Title** and **Source** shown below.

Fig 9 Default Browsing Classifiers



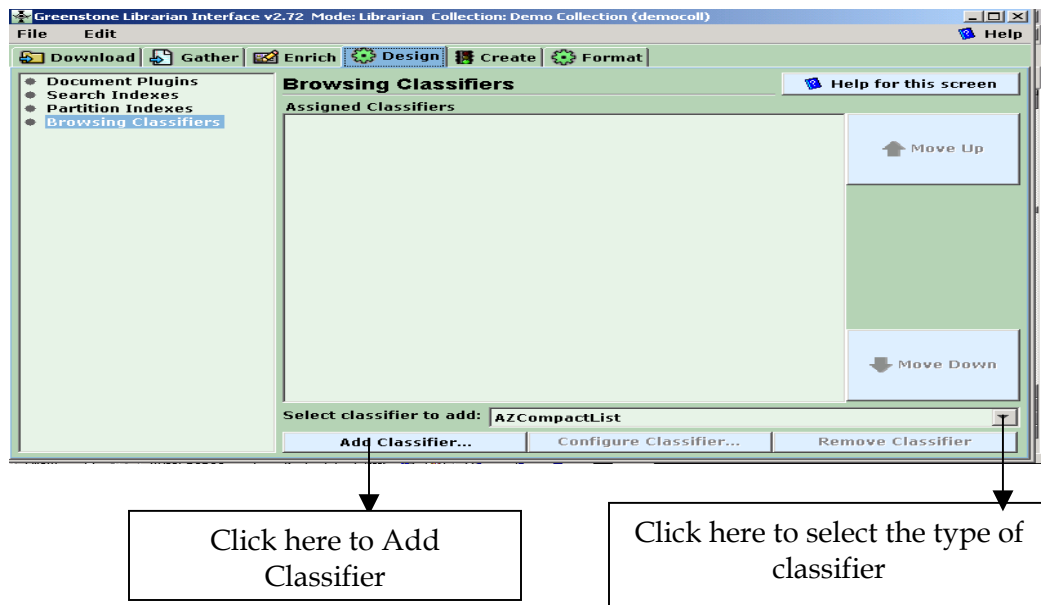
Now, choose **Select classifier to add** pull down list and select **A-Z List** or **A-Z Compact List**⁶.

Then click on **Add Classifier** and add the Browsing Classifiers⁷ for **Title**, **Creators** and **Subjects** by one by one.

⁶ Use of AZCompactList Classifier brings bookshelf icons under the browsing classifier. It groups together the documents that appear multiple times with same metadata and does not differ with AZList in any other manner.

⁷ Title, Creators and Subjects are the default button names for dc.Title, dc.Creator and dc.Subject and Keywords. You can change the *button name* by marking the *button name* check box and typing the new *button name* against it, while configuring the *browsing classifiers*. It is better to use the default *button name* initially, so as to avoid the complexities.

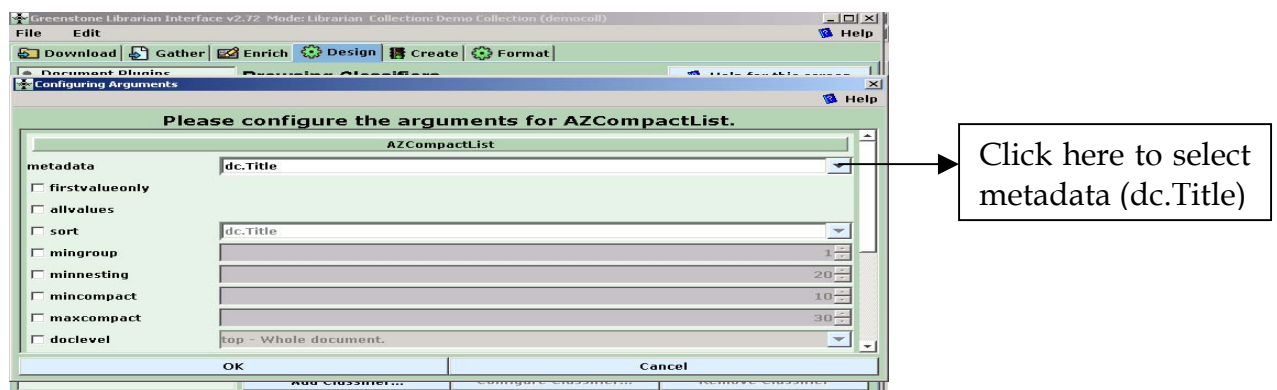
Figure 10 Adding Classifier



When you click **Add Classifier** button in the above screen, you will get the window for choosing the **Browsing Classifier**.

Select the browsing classifier for *Title* by choosing the metadata (dc.Title) as follows:-

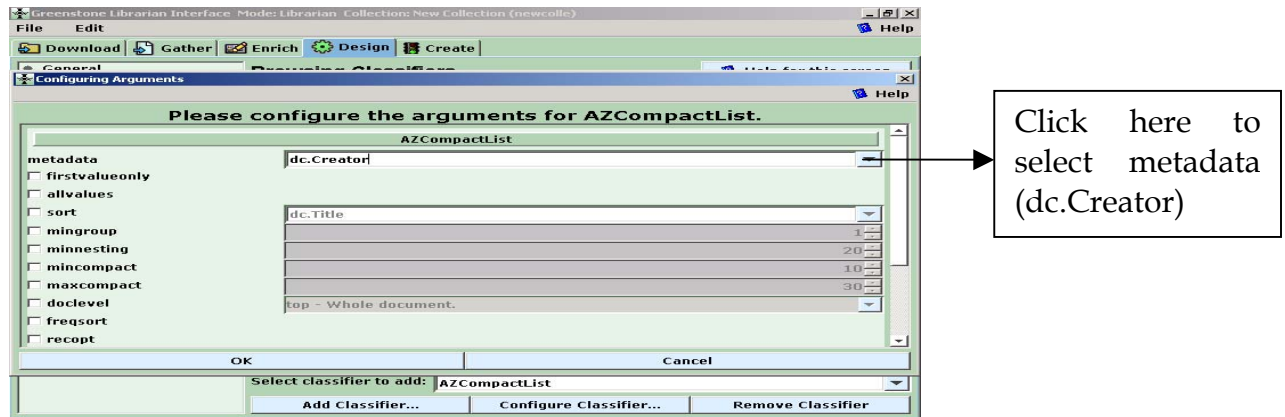
Figure 11 Adding Classifier CL1 for Title - dc.Title



Click **OK** button to add *Title* classifier.

Then select the browsing classifier for *Creator* by choosing the metadata option (dc.Creator) as follows :-

Figure 12 Adding Classifier CL2 for Creator - dc.Creator



Click OK button to add *Creator* classifier.

Then select the browsing classifier for *Subjects* by choosing the metadata option (dc.Subject and Keywords) as follows :-

Figure 13 Adding Classifier CL3 for Subjects - dc.Subject and Keywords



Click OK button to add *Subject and Keywords* classifier.

The resultant screen that shows the three added Browsing Classifiers will appear as follows:-

Fig 14 Browsing Classifiers assigned for the Demo Collection



2.4. Format

The page-display of the resultant digital library including the display-page that appears on clicking the browsing classifiers or on making a search, are governed by the features provided below the **Format** tab.

2.4.1 General

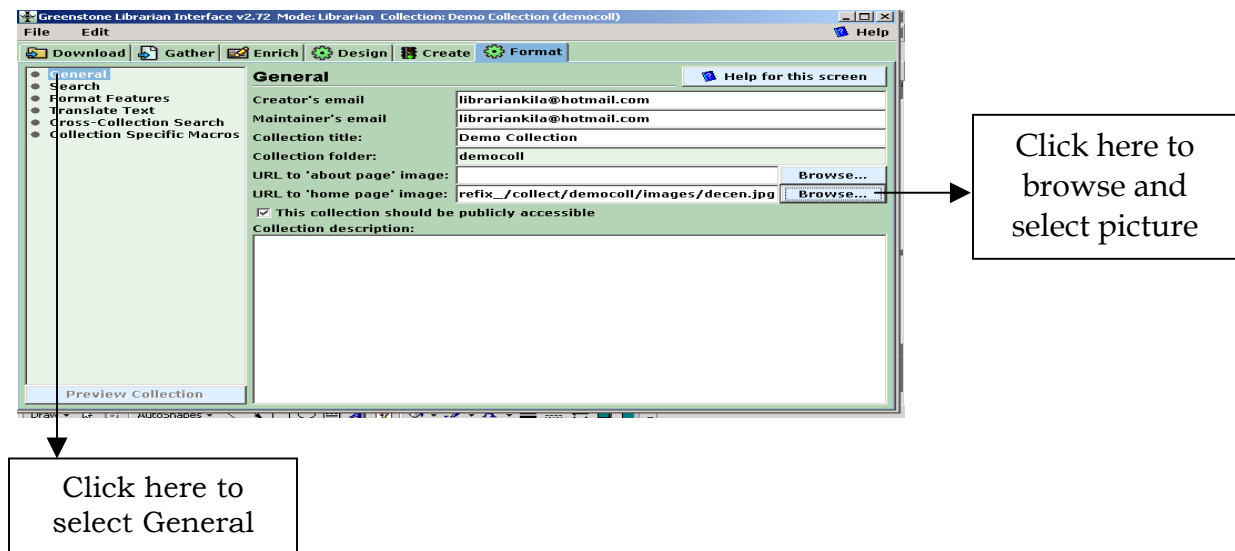
Choose the **Format** Tab and select **General** to provide the general information about the collection.

Put a picture as a collection icon

Choose a small picture of around 100*100 pixels that needs to appear as an icon of the collection in the homepage, by clicking on the **Browse** button on middle right and selecting the picture. On choosing the picture, the fullpath of the

picture will appear in the address box lying against the **Browse** button. You can provide the same picture or a different picture as the image for the 'about page' also. About page is the first page about an individual digital library collection that contains a short description about the collection.

Figure 15 Designing the Collection – General Information



2.4.2 Format Features

Beginner's may skip the *Format Features* section and go to the next section (2.5 **Build the collection**). If you skip, the default settings will take care of the page-display.

Advanced users can use the **Format Features**⁸ section of the **Format** panel on the left pane as follows: -

⁸ There are two types of Format Statements - General statement applicable in general to all and specific statement applicable to specific classifier or search list. Specific statements, if present, would override the general statements.

Let us examine some statements

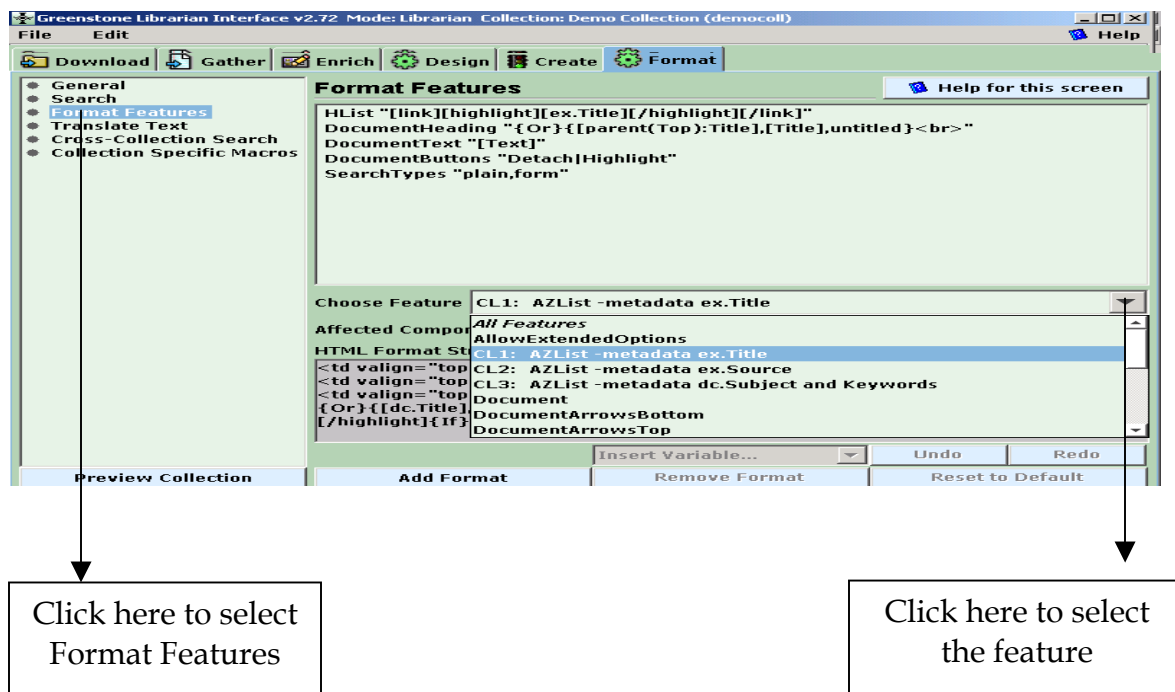
VList - applies to all vertical lists in all classifiers

Search VList - applies to all search result lists

CL2HList - applies to all hierarchy lists in classifier 2

The value of any metadata can be interpolated by placing the metadata name in square brackets. eg. [Title] in the format statements.

Fig 16 Modifying the Format Features



If you want to change any format feature, select the appropriate one (eg. CL1 AZList- metadata dc.Title) from the **Choose Feature** pull down list and add the format string to the **Format Features** by clicking on the **Add Format** button. You can select any feature appropriate for changing the appearance of the digital library. Fresh users may find it difficult to learn the format features in the beginning stage. But if properly understood, it is so easy to manage. The CL1 Browsing classifier⁹ for Title can be added as shown in the following screen.

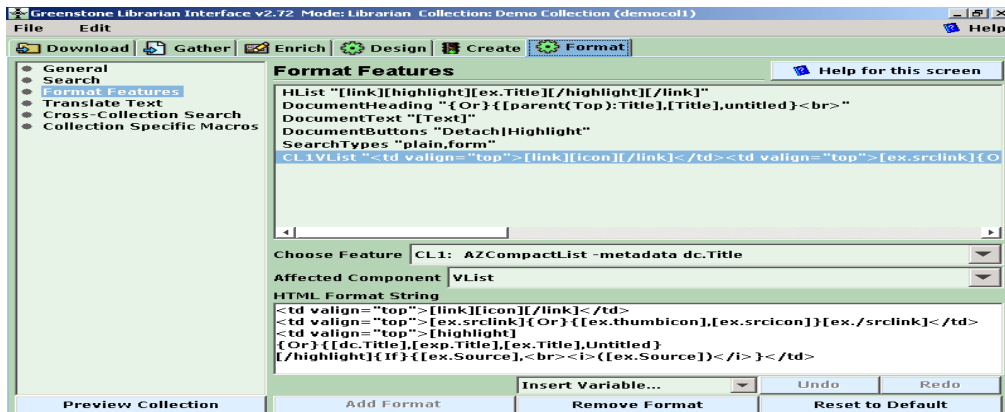
[link] [icon] [/link] provides a link from icon to the HTML version of the document.

[link] [Title] [/link] provides link from the Title of the document.

[srclink] [srcicon] [/srclink] provides a link from the word/pdf icon to the word/pdf version of the document.

⁹ CL1 denotes the first browsing classifier dc.Title and CL2 denotes the second one dc.Creator and so on.

Figure 17 Adding CL1 Browsing Classifier for dc.Title



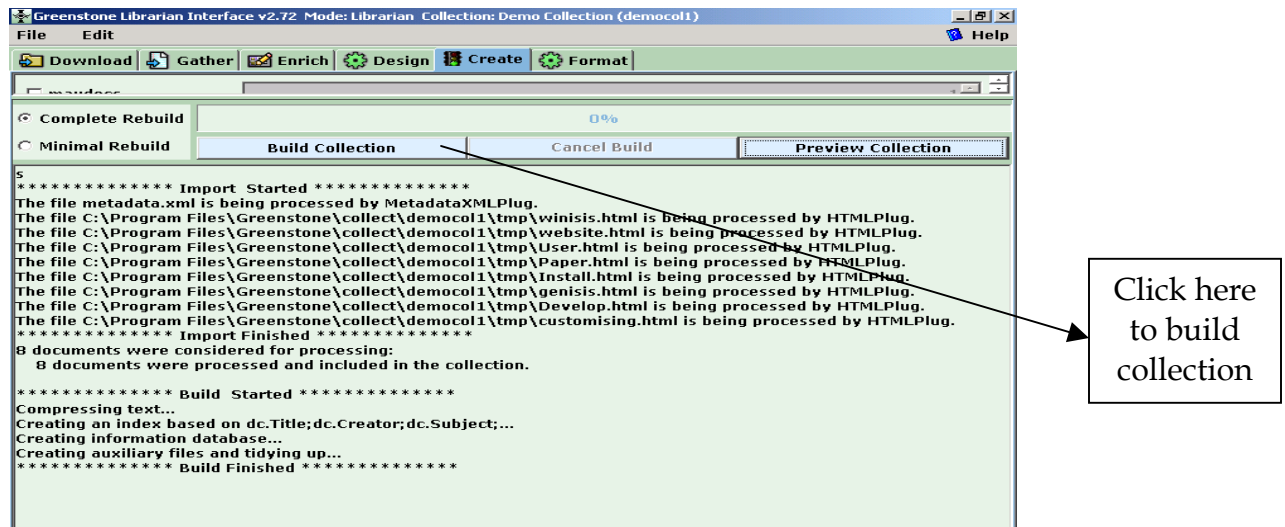
Similarly select any other string from **Choose Feature** box and click **Add Format** button for customizing it as in the above screen. You can edit the HTML strings under the Format Features box by selecting it and editing them in the **HTML Format String** box below.

You can add **CL2** and **CL3** classifiers for **Creators** and **Subjects** and can make modifications if you have reasonable knowledge of HTML.

2.5. Build the Collection

Then go to the **Create** panel. Click on the **Build Collection** button and the progress bar will show the progress in building¹⁰ the collection.

Figure 18 Building a Collection



2.5.1 Preview the Collection

At the end of the building process, Click on the **Preview Collection** button to view the collection built.

¹⁰ During the building process, the text of the documents will be compressed and the indexes specified will be generated in order to ensure the availability of the collection for search and retrieval. *GNU Data Base Manager*, a program used within the Greenstone software, will store the metadata of each document. *Managing Gigabytes* is the program used by the Greenstone system for fulfilling indexing that incorporates compression techniques.

Figure 19 Home Page of the Demo Collection

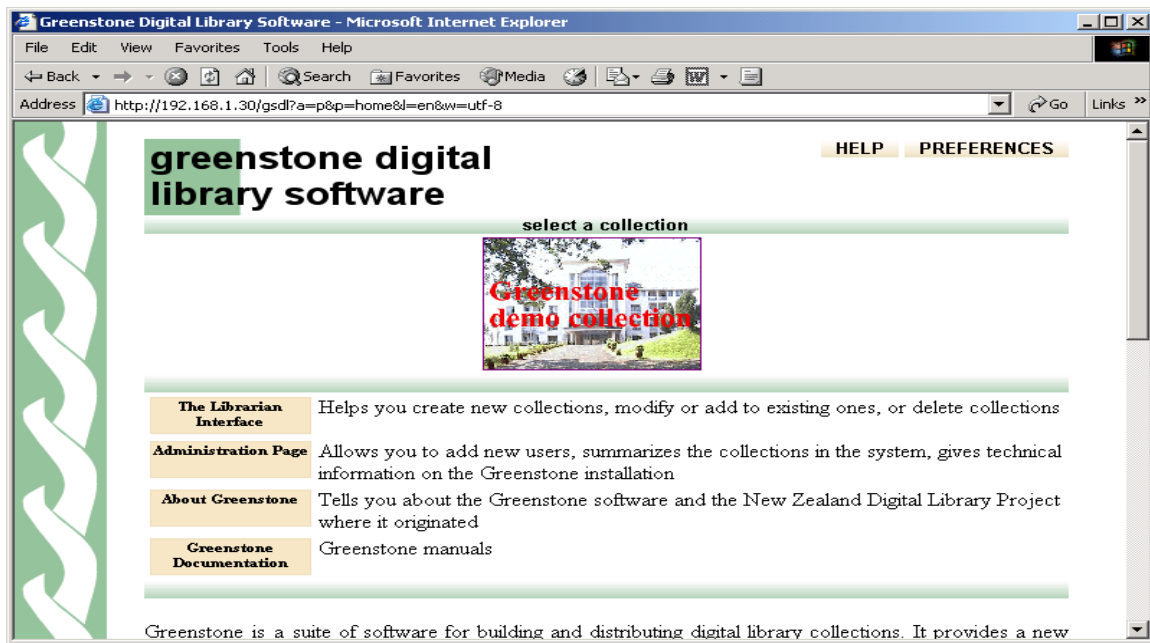


Fig 20 Browse the Collection by Title Classifier

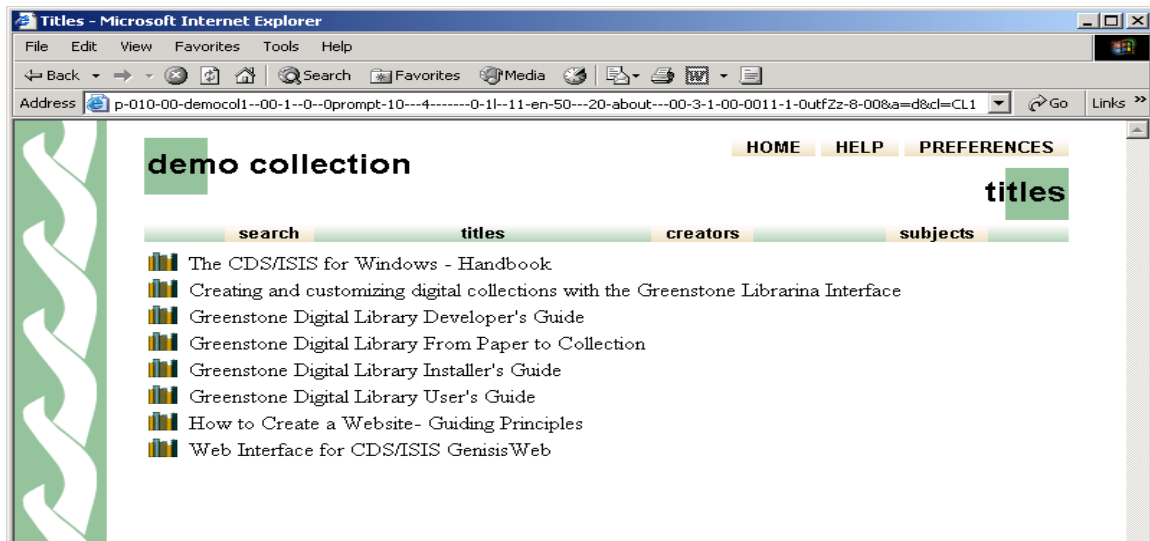
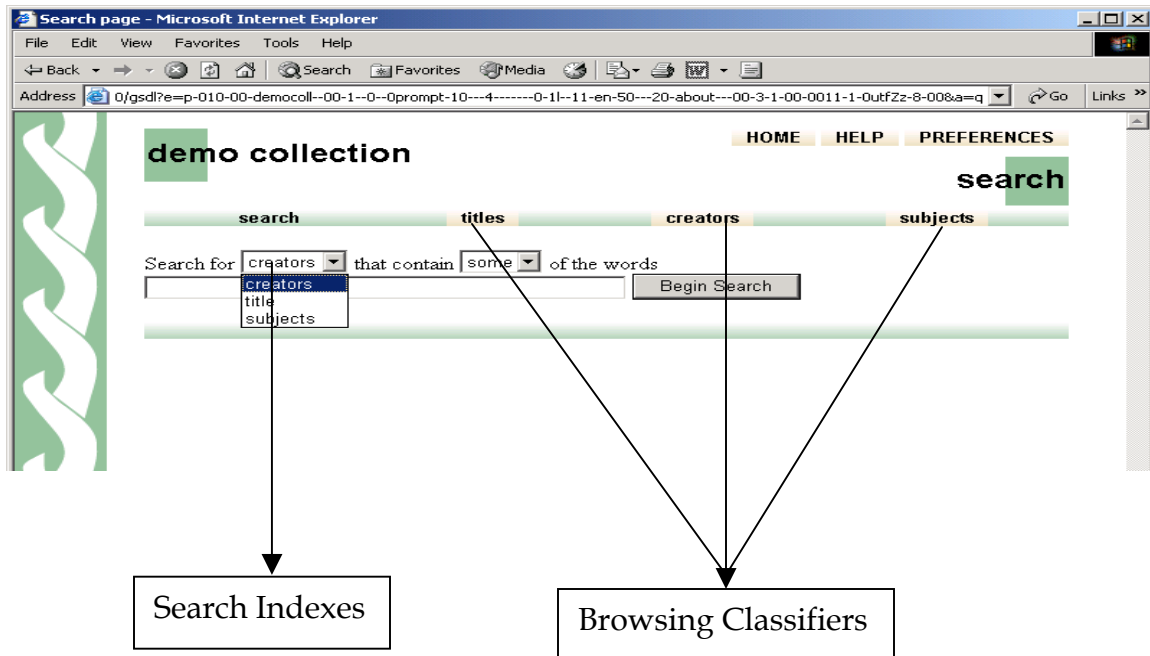


Fig 21 Search Indexes and Browsing Classifiers

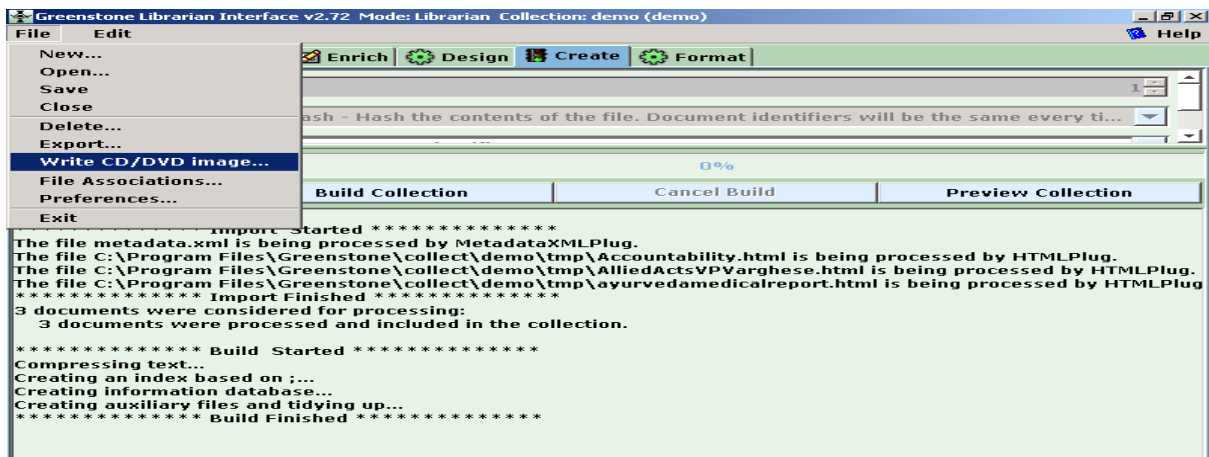


2.6. Converting the Collection in to a CD-ROM

You can export the collection/collections to a CD-ROM. If you want to convert your Greenstone application into an installable CD-ROM for distribution among wider audience.

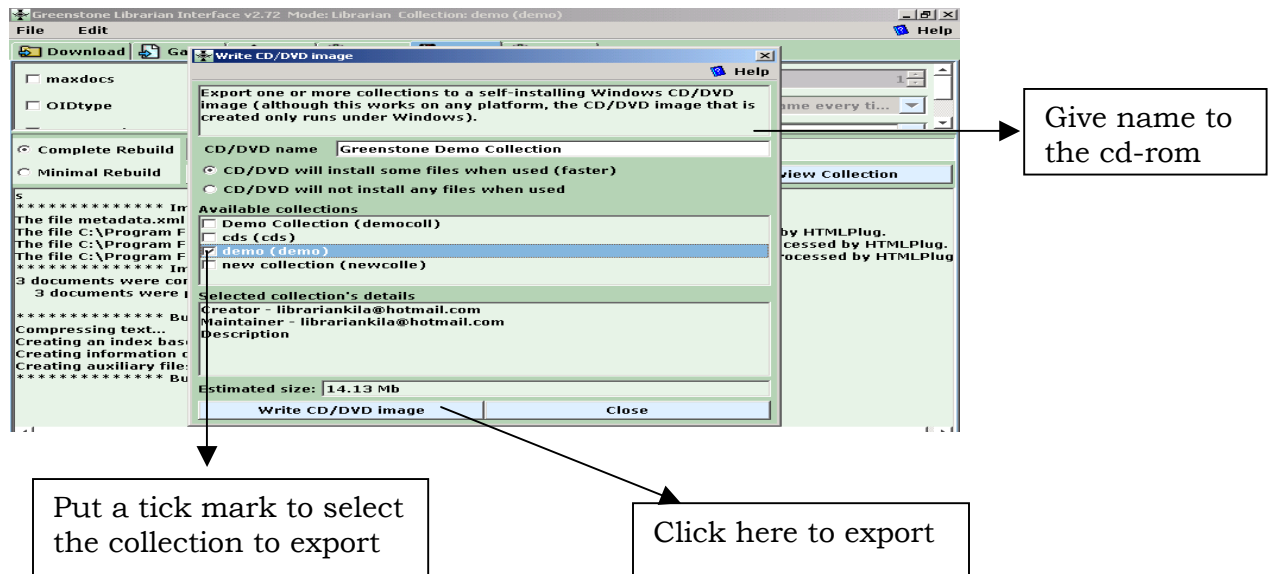
Click on File ⇒ Write CD/DVD image.....

Figure 22 Starting the Export Process



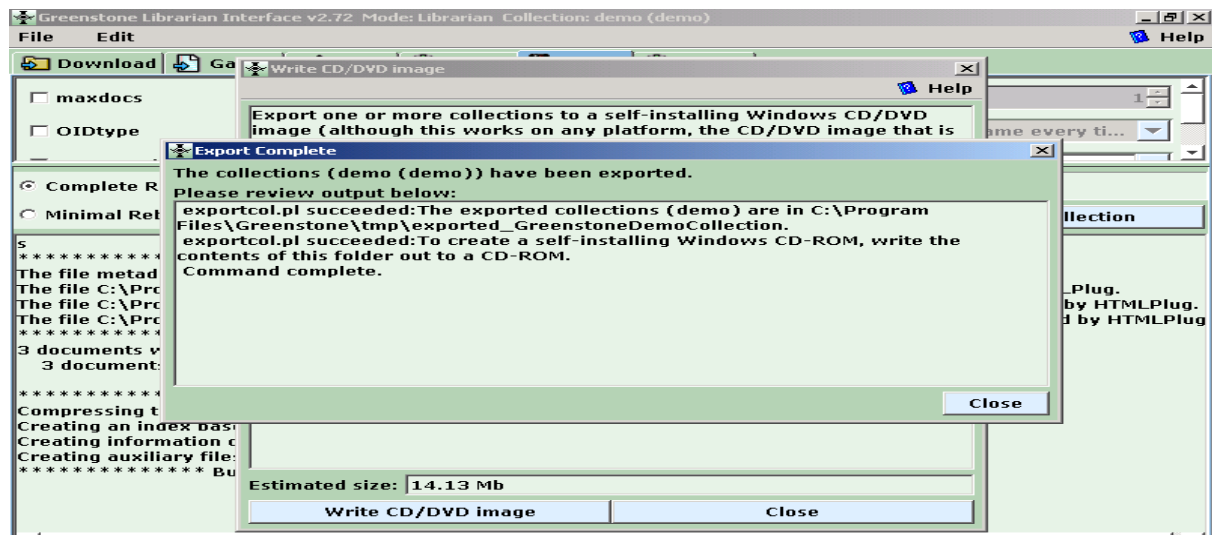
A pop-up window will appear as follows:-

Figure 23 Giving a Collection Name for CD-ROM



Provide a name for your CD-ROM, mark the check box pertaining to the collection to be exported and click *Write CD/DVD image* button.

Figure 24 Export completed



Click Close button.

Then write the contents of the folder (C:\ProgramFiles\Greenstone\tmp\exported_Demo) into a blank CD-ROM for creating self-installing Windows CD-ROM.

3. Conclusion

The basic greenstone collection with normal *look and feel* can be created within a few minutes time. The digital library collection created thus can be customized later in a variety of ways. The customization can be done, whenever you like, by opening the collection, which can be done by clicking on the menu, **File ⇒ Open** and selecting the appropriate collection in the GLI .

Format Features allow unlimited choices in modifying the appearance of the collection. You can radically alter the style of the pages generated, by changing the *macro* files included in the 'macro' folder of the Greenstone installation.

A general-purpose digital library like greenstone is a useful tool to provide information services in our libraries. Absence of knowledge on how to use it, should not come in the way of exploiting the advantages it offers. This documentation may be used as a tool to bring in more people to the growing constituency of greenstone users. We, the librarians, can improve our capabilities, as knowledge managers, if we are particular in learning the information technology tools like greenstone and use them for managing knowledge resources. We should learn, utilise, promote and propagate greenstone to make our libraries better.

References

1. Witten, Ian H and Brainbridge, David (2003) : *How to build a digital library*. Amsterdam, Morgan Kaufmann
2. Witten, Ian H and Boddie, Stefan (2004): *Greenstone Installer's Guide*. New Zealand Digital Library Project, New Zealand
3. Brainbridge, David *et al* (2004): *Greenstone digital library developer's guide*. New Zealand Digital Library Project, New Zealand
4. Loots, M. *et al* (2001) *From Paper to Collection*. New Zealand Digital Library Project, New Zealand.
5. Witten, Ian.H. *et al* (2006) *Greenstone User's Guide*. New Zealand Digital Library Project, New Zealand.
6. Morete, Pablo and Rose, John: Creating Digital Libraries Based on CDS/ISIS Databases. (*downloaded from World Wide Web*)
7. Tutorial exercises on Greenstone at
http://greenstone.sourceforge.net/wiki/index.php/Tutorial_exercises
